// 90

STAT 2509A
Test#1
SOLUTION

1. An experiment was conducted to examine the weight gain for chickens treated with various doses of a growth promotant. From a group of 15 chickens of approximately the same age and weight, a random assignment of three chickens was made to each dose group. The specified dose (mg/kg) of growth promotant was added to the feed daily for a fixed period of time. Weight gains (in pounds) are shown below:

Dose	0	0	0	0.2	0.2	0.2	0.4	0.4	0.4
Weight	1.5	1.8	1.7	2.3	2.0	1.8	4.3	3.7	4.1
Dose	0.8	0.8	0.8	1.6	1.6	1.6			
Weight	5.7	5.9	6.2	7.9	7.7	7.5			

$\sum y_i$ = 64.1 $\qquad$ $\sum x_i$ = 9

$\sum y_i^2$ = 353.59 $\qquad$ $\sum x_i^2$ = 10.2

$\sum x_i y_i$ = 57.26

[1] (a) The response variable, y, is: **weight gain of chickens** 1

[1] (b) The explanatory variable, x, is: **growth promotant dose** 1

[6] (c) State a SLR model making sure you give all assumptions necessary for statistical inference.

Model: $y = \beta_0 + \beta_1 x + \varepsilon$, ***n* = 15** 1

Assumptions: (i) *x*'s are observed without error 1

(ii) *y*'s (or ε's) are independently distributed with mean $E(y) = \beta_0 + \beta_1 x$ 1 1

(or $E(\varepsilon) = 0$**)**

(iii) variance of *y*'s (or ε's) is constant, σ^2 **for all *x*'s** 1

(iv) *y* ~ $N(E(y), \sigma^2)$ **for any value of *x* (or** $\varepsilon \sim N(0, \sigma^2)$**) for any value of *x*)** 1

[5] (d) Find the least squares estimates of β_0 and β_1. Find the least squares fitted regression line.

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} = \frac{57.26 - \frac{(9)(64.1)}{15}}{10.2 - \frac{(9)^2}{15}} = \frac{18.8}{4.8} = \mathbf{3.916667} \cong \underline{\mathbf{3.917}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{\sum_{i=1}^{n} y_i}{n} - \hat{\beta}_1 \left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{64.1}{15} - (3.916667)\left(\frac{9}{15}\right) = \mathbf{4.273333 - 3.916667(0.6) =}$$

$$\mathbf{= 1.923333} \cong \underline{\mathbf{1.923}}$$

$\therefore$ **the least squares fitted regression line is given by:** $\hat{y} = \underline{\mathbf{1.923 + 3.917}\, x}$

Assuming no violations of the assumptions, answer the following questions:

[6] (e) Find s^2, an estimate of σ^2.

$$s^2 = \frac{SSE}{n-2} = \frac{S_{yy} - \frac{S_{xy}^2}{S_{xx}}}{n-2} = \frac{\left[\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right] - \frac{\left[\sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}\right]^2}{\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right]}}{n-2} =$$

$$= \frac{\left[353.59 - \frac{(64.1)^2}{15}\right] - \frac{(18.8)^2}{4.8}}{13} = \frac{79.66933 - 73.63333}{13} = \frac{6.036}{13} = \underline{\mathbf{0.464308}}$$

$$\therefore s = \sqrt{s^2} = \mathbf{0.681401}$$

[6] (f) Use the t-test to test whether there is a significant linear relationship between the growth promotant dose and the weight gain of a chicken. Use α = 0.10.

$$H_0 : \beta_1 = 0 \qquad \alpha = 0.10 \Rightarrow \alpha/2 = 0.05$$
$$H_a : \beta_1 \neq 0$$

test-statistics: $t = \dfrac{\hat{\beta}_1}{s/\sqrt{S_{xx}}} = \dfrac{3.916667}{0.681401/\sqrt{4.8}} = \underline{\underline{\mathbf{12.59315}}}$

<u>R.R:</u> we reject H_0 if $t < -t_{\alpha/2;n-2} = -t_{0.05;13} =$ -1.771

or $t > t_{\alpha/2;n-2} = t_{0.05;13} =$ 1.771

Since t = 12.59315 > 1.771, we reject H_0 and conclude that at 10% level of significance there is an evidence to say that the growth promotant dose and weight gain of chickens are linearly related.

[4] (g) Find a 90% confidence interval for the true population slope, β_1.

$$1 - \alpha = 0.90 \Rightarrow \alpha = 0.10 \Rightarrow \alpha/2 = 0.05$$

$$\beta_1 \in \left(\hat{\beta}_1 \pm t_{\alpha/2;n-2}\, {s}/{\sqrt{S_{xx}}}\right) = \left(3.916667 \pm t_{0.05;13}\, {0.681401}/{\sqrt{4.8}}\right) = (3.916667 \pm 1.771(0.311016)) =$$

$$= (3.916667 \pm 0.550809) = (3.365858, 4.467475) \cong \underline{\underline{(3.366, 4.467)}}$$

i.e. We are 90% confident that in repeated sampling the true value of the population slope would lie in the interval (3.366 , 4.467).

[23] (h) Complete the following ANOVA table and hence test whether there is a significant linear relationship between the growth promotant dose and the weight gain of a chicken. Use α = 0.10.

$TSS = S_{yy} = \underline{\underline{\mathbf{79.66933}}}$ **(given; also calculated in part (e))**

$SSE = \underline{\underline{\mathbf{6.036}}}$ **(calculated in part (e))**

$SSR = TSS - SSE = \dfrac{S_{xy}^2}{S_{xx}} = \underline{\underline{\mathbf{73.63333}}}$ **(also calculated in part (e))**

$MSR = \dfrac{SSR}{1} = \underline{\underline{\mathbf{73.63333}}}$

$$MSE = \frac{SSE}{n-2} = \frac{6.036}{13} = \underline{\mathbf{0.464308}}$$ **(=s^2, calculated in part (e))**

$$F = \frac{MSR}{MSE} = \underline{\mathbf{158.5874}}$$

Source	d.f.	SS	MS	F
Regression	1	73.63333	73.63333	158.5874
Error	13	6.036	0.464308	
Total	14	79.66933		

$$H_0 : \beta_1 = 0 \qquad \alpha = 0.10$$
$$H_a : \beta_1 \neq 0$$

test-statistics: $F = \frac{MSR}{MSE} = \underline{\mathbf{158.5874}}$

<u>R.R:</u> we reject H_0 if $F > F_{\alpha(1,n-2)} = F_{0.10(1,13)} =$ 3.14

Since F = 158.5874 > 3.14, we reject H_0 and conclude that at 10% level of significance there is an evidence to say that the growth promotant dose and weight gain of chickens are linearly related.

[5] (i) Find the values of the coefficient of correlation, r, and coefficient of determination, r^2, and interpret their meanings in this problem. What is your conclusion about the model?

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{18.8}{\sqrt{(4.8)(79.66933)}} = \mathbf{0.961372} \cong \underline{\mathbf{96.14\%}}$$

i.e. the growth promotant dose and weight gain of chickens are strongly positively correlated (related) with the strength of their relationship of 96.14%.

$$r^2 = \frac{SSR}{TSS} = \mathbf{0.924237} \cong \underline{\mathbf{92.42\%}}$$

i.e. approximately 92.42% of the total variation in the data is explained by the regr. line (and only 7.58% is due to error). I.e. model is a good fit.

[5] (j) Find a 95% Confidence Interval of the average weight gain of chickens with the growth promotant dose of 1.5 mg/kg.

95% C.I. for *E(y)* when x_p = 1.5:
$\hat{y}$ **= 1.923333 + 3.916667(1.5) = 7.798333 and** $1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$ 1

1

$$\therefore E(y) \in \left(\hat{y} \pm t_{\alpha/2;n-2} s \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{S_{xx}}} \right) = \left(7.798333 \pm t_{0.05;13} (0.681401) \sqrt{\frac{1}{15} + \frac{(1.5-0.6)^2}{4.8}} \right) =$$

$$= (7.798333 \pm 2.160(0.3306143)) = (7.798333 \pm 0.714127) = (7.084207, 8.51246) \cong$$

1

$$\cong (7.084, 8.512)$$ 1

i.e. We are 95% confident that in repeated sampling the average weight gain of chickens who were administered a growth promotant dose of 1.5 mg/kg would be between 7.084 and 8.512 lbs. 1

[5] (k) Find a 95% Prediction Interval of the weight gain of a chicken who was administered the growth promotant dose of 1.5 mg/kg.

95% P.I. for *y* when x_p = 1.5:
$\hat{y}$ **= 1.923333 + 3.916667(1.5) = 7.798333 and** $1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$ 1

1

$$\therefore y \in \left(\hat{y} \pm t_{\alpha/2;n-2} s \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{S_{xx}}} \right) = \left(7.798333 \pm t_{0.05;13} (0.681401) \sqrt{1 + \frac{1}{15} + \frac{(1.5-0.6)^2}{4.8}} \right) =$$

$$= (7.798333 \pm 2.160(0.75737253)) = (7.798333 \pm 1.635925) = (6.162408, 9.434258) \cong$$

1

$$\cong (6.162, 9.434)$$ 1

i.e. We are 95% confident that in repeated sampling the weight gain of a chicken who was administered a dose of growth promotant of 1.5 mg/kg would be between 6.162 and 9.434 lbs. 1

2. Refers to question 1.

Dose x_i	**Weight gain** y_{ij}	n_i	$\bar{y}_i$	$\sum_j (y_{ij} - \bar{y}_i)^2$
0	1.5 1.8 1.7	3	1.6666667	0.0466666
0.2	2.3 2.0 1.8	3	2.0333333	0.1266666
0.4	4.3 3.7 4.1	3	4.0333333	0.1866666
0.8	5.7 5.9 6.2	3	5.9333333	0.1266666
1.6	7.9 7.7 7.5	3	7.7	0.08

[5] (a) Decompose SSE into the sum of squares due to the pure error, SSPE, and sum of squares due to the lack of fit, SSLF.

Hint: $SSPE = \sum_i \sum_j (y_{ij} - \bar{y}_i)^2 = 0.5666664$

$\sum x_i = 9$ $\sum x_i^2 = 10.2$ $\sum y_i = 64.1$ $\sum y_i^2 = 353.59$ $\sum x_i y_i = 57.26$

Solution: 1

$SSE = SSPE + SSLF$ 1

$SSE = S_{yy} - \dfrac{S_{xy}^2}{S_{xx}} = \underline{\underline{6.036}}$ **(calculated in Q.1(e))**

$SSPE = \underline{\underline{0.5666664}}$ **(given)** 1

$\therefore SSLF = SSE - SSPE = \underline{\underline{5.469334}}$ 1 1

[6] (b) Test whether the linear model $y = \beta_0 + \beta_1 x + \varepsilon$ is appropriate. Use $\alpha = 0.05$.

$y = \beta_0 + \beta_1 x + \varepsilon$, $\alpha = 0.05$

H_0 : *model is appropriate*
H_a : *model is not appropriate* 1

1

test-statistics: $F = \dfrac{MSLF}{MSPE} = \dfrac{SSLF \Big/ \left[(n-2) - \sum_i (n_i - 1)\right]}{SSPE \Big/ \sum_i (n_i - 1)} = \dfrac{5.469334/(13-10)}{0.5666664/10} =$

$= \dfrac{1.823111}{0.05666664} = \underline{\underline{32.17257}}$ 1

R.R: **we reject** H_0 **if** $F > F_{\alpha(n-2-\sum_i (n_i-1), \sum_i (n_i-1))} = F_{0.05(3,10)} =$ **3.71** 1

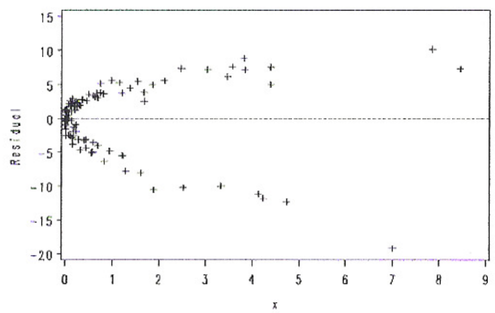
1 1/2

Since F **= 32.17257 > 3.71, we reject** H_0 **and conclude that at 5% level of significance there is enough evidence to say that a linear model is not appropriate.**

1/2

3. State which violations of the SLR model (if any) are indicated by each of the following residual plots. Give reasons for your answer.

[3] (a)

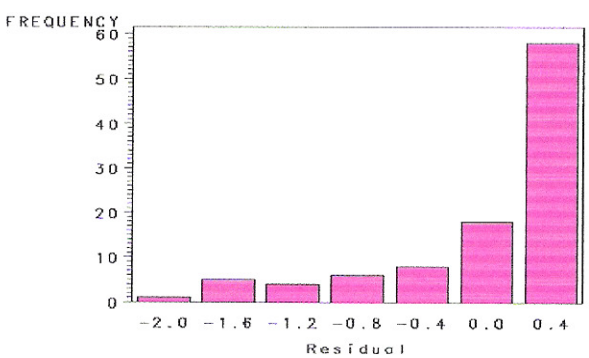


1

- **Violation of the assumption of constant variance, since the residuals are increasing with x's**

2

[3] (b)

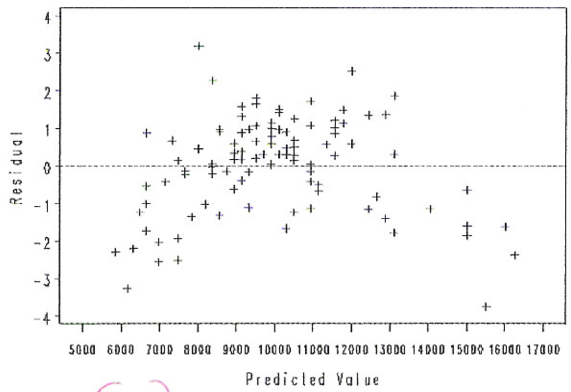


1

- **Violation of the assumption of errors being normally distributed, since the histogram of errors is not bell-shaped, nor is it symmetric (it is negatively skewed)**

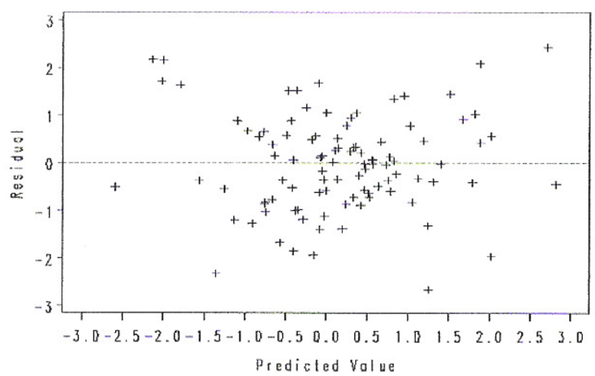
1 1

[3] (c)



- **Violation of independence (or linearity), since we have a curve-linear pattern**

[3] (d)



- **No violations, since residuals are randomly scattered around their mean (i.e. no pattern)**